

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form ([see an example](#)) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Quantifying public preferences for different bowel preparation options prior to screening CT colonography: A discrete choice experiment
AUTHORS	Ghanouni, Alex; Halligan, Steve; Taylor, Stuart; Boone, Darren; Plumb, Andrew; Stoffel, Sandro; Morris, Stephen; Yao, Guiqing Lily; Zhu, Shihua; Lilford, RJ; Wardle, Jane; von Wagner, Christian

VERSION 1 - REVIEW

REVIEWER	Patrick Bossuyt University of Amsterdam, the Netherlands.
REVIEW RETURNED	05-Dec-2013

GENERAL COMMENTS	<p>There is confusion between preference and uptake</p> <p>The authors present a discrete choice experiment (DCE) in which 607 participants expressed their preference for hypothetical CTC screening programs that differed in terms of sensitivity (86% to 92%), specificity (89% to 91%) and bowel preparation (non-laxative, mild-laxative, or powerful-laxative).</p> <p>The DCE is reported in detail. The design and the results are well described, and easy to follow.</p> <p>I have a number of concerns and invitations for clarification.</p> <p>1. Preference or choice?</p> <p>Page 7 – “These estimates can then be used to predict which method of delivering CTC would achieve the highest level of uptake if offered for screening.”</p> <p>Page 11 – “Each comparison consisted of two stages: first, participants were asked to state which of the two tests they thought appeared best, after which the less preferred option was faded out and participants were asked whether they would have the preferred test if it were offered to them in the next month (or if they would opt for no testing; ..). The second stage allowed estimates of screening uptake to be calculated and a two-step approach may improve comprehension of the various attributes”</p> <p>As the authors themselves indicate, the task is not only hypothetical, but also artificial. In an actual screening situation, most countries would offer invitations to only one population screening program. (There are exceptions) As indicated the task is presented in two stages: first eliciting a preference, and then inviting a choice (whether or not to take the preferred test).</p> <p>The second task is more valid (“would you participate in this</p>
-------------------------	--

	<p>program?”), and I am surprised that only one program is presented in the second stage. One cannot infer participation, not even intended, from preferences. It is very well possible that the participant would also willing to participate in the least preferred option, despite the existence - hypothetical or not – of a preferred alternative. Only if the participant is not willing to participate one can infer that he or she would also not take part in the least preferred screening program.</p> <p>It would have been better to ask for choices (participation) rather than preferences, so one could get a better understanding of the consequences for the effectiveness of screening programs.</p> <p>The language used in the paper is quite confusing. On page 14 the authors refer to “Estimates of uptake for all possible permutations of preparation type and sensitivity are shown in Table 6”, but Table 6 expresses the “probability of choosing the test”, which is not the same as uptake. Were these probabilities based on the answers in the first stage, or in the second? This was not clear to me; please describe how you calculated these.</p> <p>I know data collection has already been completed, so this cannot be corrected. The authors could change the wording of their manuscript, to better reflect the difficulties in projecting uptake from preferences, and to maintain consistency in describing the difference between preference and choice (uptake).</p> <p>2. Page 11 – “The primary analysis used conditional logistic regression, grouped by a unique identifier for each choice set and each participant.”</p> <p>I did not understand this sentence. How was conditioning done? Page 10 specifies that every participant evaluated only one choice set.</p> <p>3. Page 12 – “Most participants identified the greatest risk of a disease correctly, and reported that generally they found medical statistics and written medical information easy to understand.”</p> <p>I find this a very generous interpretation of the results, as presented in Table 3. Only 67% correctly identified the greatest risk of disease (objective numeracy), and one in four indicated they found understanding written medical information hard or very hard.</p> <p>4. Page 10 – “Participants were then given information about the three methods of preparation (..), followed by information regarding the practicalities of colonoscopy.”</p> <p>Page 25 – “Specificity “Out of 100 people without a polyp, how many people would get a false alarm.”</p> <p>I am somewhat concerned about the way information was presented to the participants, especially about specificity. Was the exact sentence above used in the description? A “false alarm” sounds relatively innocent, compared to “colonoscopy”.</p> <p>5. Page 6 – “CT colonography (CTC) has been recommended as a screening test for colorectal cancer (CRC). It is capable of high sensitivity for cancer and clinically significant precancerous Polyps.”</p> <p>Why not just “It has a high sensitivity (etc).”</p>
--	--

	6. The authors refer to the Dutch trial, comparing colonoscopy and CTC. The burden of both techniques was also evaluated in that trial, see "Burden of colonoscopy compared to non-cathartic CT-colonography in a colorectal cancer screening programme: randomised controlled trial" by de Wijkerslooth TR et al. in Gut. 2012 Nov;61(11):1552-9.
--	--

REVIEWER	dr. M.H. Liedenbaum AMC Amsterdam, the Netherlands
REVIEW RETURNED	15-Dec-2013

GENERAL COMMENTS	<p>-Title: The title is very long and first gives the impression that it is about a study that analyses CTC sensitivity in different bowel preps. A title such as: ' Quantifying public preferences via a discrete choice experiment in screening CT colonography ' might come more to the point.</p> <p>-Abstract: I find the first sentence in the conclusion not very comprehensible (page 5, line 3-7). A sentence at the end of the article in the conclusion might be a better alternative: Uptake of CTC following non- or reduced laxative preparations are not likely to increase uptake as any gains in perceived burden are undermined by the reduction in sensitivity.</p> <p>-Introduction: what is the situation in your country now in CRC screening? Is CTC one of the screening options? Do people have a choice in screening tests?</p> <p>-Results: the authors have gained demographic information of the patients. This is displayed in a table. My question however; what is the influence of these demographics on the experiment? E.g.: are higher educated people more willing to choose a full laxative option with a higher sensitivity and is there an influence of having had bowel testing before?</p> <p>- Discussion: the authors state that people's expectations about cancer risk can be unrealistically high. (page 15, line 25,26). Furthermore it is likely that people cannot estimate what the difference is between a sensitivity of 92% versus 86%. And even if this difference would have been larger (15% or more), people might still accept the lower sensitivity for a less burdensome bowel prep because they cannot estimate the consequences of this difference. It is a task for us as physicians to tell which sensitivity is acceptable. And because patients prefer less burdensome bowel preparations we must try to find acceptable bowel preparations with high sensitivity.</p> <p>-One of the limitations is as well that there has only been tested on preferences in one screening modality (CTC). Other realistic options such as FOBT and colonoscopy have not been addressed.</p>
-------------------------	---

VERSION 1 – AUTHOR RESPONSE

REVIEWER #1 (PATRICK BOSSUYT):

1a) In both stages of every comparison, the participant is asked to make a choice. The first choice is between two options that have different combinations of the levels of each of the attributes. The second choice is between the preferred option from the first choice and a no testing option. This is an entirely valid approach to use in a DCE and has been used previously in other studies (e.g. 1). The lead author of the previous study has also contributed to guidelines for optimal conduct of DCE research (2). The first stage is vital to include because the point of a DCE is to quantify the relative importance of the different attributes, which is what this stage does. As the reviewer has pointed out, it is also possible that the respondent may choose not to accept a test at all, which is why we have included the second stage. It is also important to note that each respondent is not given only one comparison like this, but is asked to repeat the process several times, each time comparing between choices with different levels of attributes, and then with no testing. Furthermore, most previous DCE studies of colorectal cancer screening use a single-stage design and only ask participants to state a preference for “test A”, “test B” and “no testing” (e.g. 3-4). These illustrate that there is no effective distinction between asking participants about preferences for “test A” vs. “test B” vs. “no testing” and asking participants about preferences for “test A” vs. “test B”, followed by preferences for “test A|B” vs. no testing”.

1. Marshall DA, Johnson FR, Phillips KA, Marshall JK, Thabane L, Kulin NA. Measuring patient preferences for colorectal cancer screening using a choice-format survey. *Value Health*. 2007;10(5):415–430.
2. Bridges JFP, Hauber AB, Marshall D, et al. Conjoint analysis applications in health - a checklist: A report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value Health*. 2011;14(4):403–413.
3. Hol L, de Bekker-grob, EW, van Dam L, et al. Preferences for colorectal cancer screening strategies: a discrete choice experiment. *Brit J Cancer*. 2010;102(6):972–980.
4. van Dam L, Hol L, de Bekker-grob EW, et al. What determines individuals' preferences for colorectal cancer screening programmes? A discrete choice experiment. *Eur J Cancer*, 2010. 46(1):150–159.

1b) We agree with the reviewer's point that “estimates of uptake” is not a satisfactory description. We have amended the text and table title to the more accurate “probability of selecting a test”. The probabilities of selecting a test were based on answers from both the first stage and the second stage. The first stage calculated coefficients, relating to the effect that each attribute had on preferences. This also determined which attributes should be included in the probability calculations. The second stage uses participants' actual decisions in the survey and the obtained coefficients to calculate predictions. The “predict p1” command was used in Stata to create a new variable for each possible combination of levels based on the statistically significant coefficients (preparation and sensitivity). Hence, the command “summ p1 if false_negative ==float(.92) & laxative ==3 & neither ==0” gives the predicted probability that individuals choose a screening test that has 92% sensitivity and full-laxative preparation.

We have attempted to clarify this in the penultimate paragraph of the “Analysis” section.

2) Thank you for drawing our attention to the point on Page 10, which was a typing error – this should have referred to choice scenarios, not choice sets. Each participant therefore had 3-4 IDs which were different to all other participants' IDs. We have aimed to clarify how conditioning was done in the second paragraph of the “Analysis” section.

3) We agree with that reviewer that this is a relevant caveat and have been more explicit in the "Results" section, paragraph 2.

4) In addition to referring to "false alarms" (indicating the potential for unwarranted worry regarding the diagnosis), we also provided information indicating that undergoing (unnecessary) colonoscopy was associated with inconvenience and discomfort. We provided information about what the test involved along with the concept of "false alarms" to indicate that colonoscopy and false alarms were closely related (Appendix 2, pages 7-8).

5) We have made this correction in paragraph 1 of the Introduction.

6) We agree that there is scope to relate the findings of this study with our own and have referenced it in paragraph 2 of the introduction and considered it in more detail in the discussion (paragraph 2).

REVIEWER #2 (DR M.H. LIEDENBAUM):

Title: We have suggested a shortened title on the title page.

Abstract: We agree that this sentence was unnecessarily difficult to follow and have adapted the suggested sentence into the conclusion of the Abstract.

Introduction: We have described the current status of UK colorectal cancer screening in paragraph 1 of the Introduction.

Results: This study was only designed to assess preferences among the general population approaching screening age, collectively, since the UK (similar to organised screening programmes in many other countries) currently only offers a single test to the entire screening-eligible population. However, we agree that these are interesting and potentially important follow-up questions arising from the study and have elaborated on them in paragraph 6 of the Discussion.

Discussion A: As we argue in the discussion, we would agree that the onus is on clinicians and other scientists to improve burden of preparations and we also agree that it is vital that medical and healthcare professionals consider sensitivity to be acceptable enough for a given test to be offered at all. However, we also argue that it is ultimately screening invitees who make the decision as to whether a given level of sensitivity is acceptable enough for them to undergo the test. Invitees report that they do wish to know about test sensitivity (e.g. 5) so it is reasonable to seek their views, which may reveal that invitees would consider some levels of sensitivity to be unacceptably poor and simply not attend. Likewise, very high levels of sensitivity are likely to be more acceptable to invitees, which may promote higher levels of uptake and improve population health outcomes.

5. Ling BS, Moskowitz MA, Wachs D, Pearson B, Schroy PC, & Schroy III PC. Attitudes toward colorectal cancer screening tests. A survey of patients and physicians. *J Gen Intern Med*, 2001. 16(12):822–830.

Discussion B: We agree that the study could have explored preferences for other investigations, as others have done. However, we decided to narrow the scope of the study by prioritising a simpler DCE that participants would find less confusing, with the intention of answering a more specific research question.

VERSION 2 – REVIEW

REVIEWER	Patrick Bossuyt University of Amsterdam, the netherlands
REVIEW RETURNED	06-Feb-2014

GENERAL COMMENTS	<p>The authors have made a number of changes, in response to the reviewers' comments. These changes helped to clarify the project, with conclusions that are more in line with their study design and findings.</p> <p>For me, it also generated new areas of confusion, which I feel now needs clarification.</p> <p>1. Analysis</p> <p>As explained under "Measures" the task consisted of a series of paired comparisons. After each comparison participants were asked whether they "would have the preferred test if offered to them in the next month".</p> <p>The first task invites a choice, based on preference ("A or B"). The second task asks for an intended decision about uptake "would you take A: yes/no" (if A was the preferred option, B otherwise). The second task</p> <p>The analysis section contains the following lines "The primary analysis used conditional logistic regression (...) The model incorporated the three effects-coded laxative variables, plus centred sensitivity and specificity variables. Significant coefficients ($p < 0.05$) denote that an attribute was associated with preferences."</p> <p>The term "preferences" is confusing here, introducing ambiguity.</p> <p>Did the authors analyse the first part of each task (the preference part: A or B), or the second part (would you take A: yes/no).</p> <p>It is possible to analyse pairwise comparisons using logistic regression (using the difference in linear components to explain the probability of choosing A rather than B). It is also possible to analyze the second part (the decision – A: yes/no) with logistic regression. The wording seems to suggest that the later was done. If so, please clarify.</p> <p>If the authors analyzed intended uptake, what then was done with the preference data (A or B)? How were they analyzed? Or were they not analyzed at all, and just introduced to "improve comprehension of the attributes" (pages 9-10). In that case, preferences were not used at all in the data analysis, and all conclusions about "preferences" should be removed from the manuscript (see also final paragraph in the discussion section).</p> <p>Please clarify the analysis section, and use consistent language to distinguish preferences from intended uptake. This also applies to the final paragraph of the discussion.</p> <p>Other remarks</p>
-------------------------	---

	<p>2. Page 5 – “guaiac faecal occult blood testing (...) has little potential for prevention.”</p> <p>"Little potential? " This is definitely not the general consensus. See, for example, the commentary of Graeme Young in Gut, 2012. Please rephrase, or quote Scholefield and colleagues literally.</p>
--	---

VERSION 2 – AUTHOR RESPONSE

R1: Analysis: As explained under “Measures” the task consisted of a series of paired comparisons. After each comparison participants were asked whether they “would have the preferred test if offered to them in the next month”. The first task invites a choice, based on preference (“A or B”). The second task asks for an intended decision about uptake “would you take A: yes/no” (if A was the preferred option, B otherwise). The analysis section contains the following lines:

"The primary analysis used conditional logistic regression...The model incorporated the three effects-coded laxative variables, plus centred sensitivity and specificity variables. Significant coefficients ($p < 0.05$) denote that an attribute was associated with preferences."

The term “preferences” is confusing here, introducing ambiguity. Did the authors analyse the first part of each task (the preference part: A or B), or the second part (would you take A: yes/no)?

It is possible to analyse pairwise comparisons using logistic regression (using the difference in linear components to explain the probability of choosing A rather than B). It is also possible to analyze the second part (the decision – A: yes/no) with logistic regression. The wording seems to suggest that the later was done. If so, please clarify.

A: “We have aimed to clarify that in each scenario, responses across both stages are combined to determine stated overall preferences for one of three options: having Scan A, having Scan B, having no testing. Hence, we have not distinguished between preferences and intended uptake but clarified how both parts of the choice scenarios correspond to overall preferences in the third paragraph of the “measures” section”:

Overall preferences for each choice scenario were determined over two stages: first, participants were asked to state which of the two tests they thought appeared best (i.e. which of the two tests they preferred), after which the non-preferred option was faded out and participants were asked whether they would have the preferred test if it were offered to them in the next month or if they would opt for no testing (i.e. whether their overall preference was for having the initially favoured test or having no testing; Figure 1. for an example).

“Taking this into account, we have also aimed to clarify the outcome of the logistic regression in the second paragraph of the analysis section. Each choice scenario for each participant generated three interrelated observations: prefers to have Scan A overall; prefers to have Scan B overall and prefers to have no testing overall”:

The first stage of the analysis used conditional logistic regression with three effects-coded laxative variables, plus centred sensitivity and specificity variables as predictors. The outcome was whether a given option was preferred overall i.e. each choice scenario for each participant generated three observations: Scan A preferred overall (yes/no), Scan B preferred overall (yes/no) or no testing preferred overall (yes/no). A unique identifier was generated to account for the interrelated nature of responses within choice sets and participants (i.e. four participants completing three choice sets had

twelve unique identifiers between them).

R1: If the authors analyzed intended uptake, what then was done with the preference data (A or B)? How were they analyzed? Or were they not analyzed at all, and just introduced to “improve comprehension of the attributes” (pages 9-10). In that case, preferences were not used at all in the data analysis, and all conclusions about “preferences” should be removed from the manuscript (see also final paragraph in the discussion section).

A: “As noted above, the A vs. B comparison represented the first stage of establishing overall preferences for each scenario and so were included in the estimated probabilities of choosing a test, which served as a proxy of intended uptake. We have amended the third paragraph of the analysis section to clarify this.”

Finally, we calculated the probability that a test would be selected (using the “predict p1” and “summ p1 if...” commands in Stata). This extrapolated the statistically significant coefficients observed in the primary analysis to create a new variable for each of the possible combinations of levels for the statistically significant attributes. This was calculated for all choice scenarios where a participant had an overall preference for either Scan A or Scan B (i.e. did not select no testing).

R1: Please clarify the analysis section, and use consistent language to distinguish preferences from intended uptake. This also applies to the final paragraph of the discussion.

A: “As we have outlined, we have attempted to clarify that we are referring to intended uptake as a product of participants’ preferences: specifically, for having the favoured scan over the alternatives of having the less favoured scan and having no testing at all.”

R1: Page 5 – “guaiac faecal occult blood testing (...) has little potential for prevention.”

"Little potential? " This is definitely not the general consensus. See, for example, the commentary of Graeme Young in Gut, 2012. Please rephrase, or quote Scholefield and colleagues literally.

A: “We acknowledge that “little potential” in the introduction is too strong and have noted instead that guaiac FOBt has yet to demonstrate CRC prevention using the UK protocol and at UK levels of uptake, which is more consistent with the Young et al. commentary.”

CT colonography (CTC) has been recommended as a screening test for colorectal cancer (CRC).^{1,2} It is capable of high sensitivity,^{3,4} potentially reducing CRC mortality and incidence.⁵ In this respect, it may represent an improvement upon the only widely available method of screening in the UK (guaiac faecal occult blood testing), which has yet to demonstrate preventative potential with an uptake level of 57% and a threshold for positivity of 5-6 abnormal samples out of a possible six.⁶

VERSION 3 - REVIEW

REVIEWER	Patrick Bossuyt University of Amsterdam
REVIEW RETURNED	12-Mar-2014

- The reviewer completed the checklist but made no further comments.